DOCUMENT RESUME

ED 207 988 TM 810 569

TITLE	1981 Local Option Testing Manual. Wisconsin Pupil Assessment Program.
INSTITUTION	Wisconsin State Dept. of Public Instruction, Madison. Div. for Management and Planning Services.
REPORT NO	WSDPI-Bull-1192
PUB DATE	81
NOTE	33p.
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Academic Standards; *Criterion Referenced Tests; Economics; *Elementary Secondary Education; Mathematics; Norm Referenced Tests; *Performance; Reading; Sampling; Scores; *State Programs; *Testing Programs; Writing (Composition)
IDENTIFIERS	*Wisconsin Pupil Assessment Program

ABSTRACT

The Wisconsin Pupil Assessment Program is a testing service which provides the State of Wisconsin with: (1) a measure of student performance in selected academic areas by developing objective-referenced tests in reading, mathematics, economic understanding, and writing; (2) a comparison of Wisconsin public school students to the rest of the nation in mathematics, reading, and language by administering a standardized norm-referenced test; and (3) a description of student scores over a period of years by reporting annually trends in performance on the above-mentioned tests. Further, the service allows public and private schools who voluntarily use its products and processes to administer the objective-referenced tests used for purposes of statewide assessment. An overview of these statewide tests is provided and the test objectives in mathematics, reading, writing, and economic understanding are listed. All tests are identified with specific grade levels, but they can be administered at other levels. A general discussion on whether to test all students or select a random sample is given. Guidelines for the setting of performance standards for test objectives are provided and five methods which might be used are described. (AL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED207988

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

1981 LOCAL OPTION TESTING MANUAL

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G. Carter

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Wisconsin Pupil Assessment Program

TM 810 669

Wisconsin Department of Public Instruction / Barbara Thompson, Ph.D., State Superintendent

1981 LOCAL OPTION TESTING MANUAL

DIVISION FOR MANAGEMENT,
PLANNING, AND FEDERAL SERVICES

A. A. Buchmiller, Assistant Superintendent

Bureau for Evaluation, Planning, Assessment,
and Research
Tom Stefonek, Director

Wisconsin Pupil Assessment Program
Russell Allen
Diedre Brensike
Maggie Burke
Vicki Fredrick
Darwin Kaufman
Susan Ketchum
Angie Wisinger

Bulletin No. 1192

Table of Contents

	Page
Introduction	v
Overview of Statewide Tests	1
Test Objectives in	
Mathematics	4
Geometry	6
Reading	8
Writing	12
Economic Understanding	13
General Testing Information	
Sampling or Testing All Students	15
Setting Performance Standards	15

4/5

Introduction

The Wisconsin Pupil Assessment Program has four major goals:

1. To provide the legislature, governor's office, and citizens of Wisconsin with a measure of student performance in selected academic areas.

   To accomplish this first goal, the Pupil Assessment Program has developed objective-referenced tests in reading, mathematics, economic understanding, and writing.

2. To provide the legislature, the governor's office, and general citizenry of Wisconsin with a comparison of Wisconsin public school students to the rest of the nation in mathematics, reading, and language.

   The Pupil Assessment Program administers a standardized norm-referenced test to satisfy this goal.

3. To provide the legislature, the governor's office, and citizens of Wisconsin with a description of student scores over a period of years.

   Systematic trends in performance on the objective-referenced tests and the norm-referenced test are reported annually by the Wisconsin Pupil Assessment Program.

4. To provide materials and technical assistance to schools which voluntarily use the products and processes of the Pupil Assessment Program.

   In order to accomplish this goal, the Wisconsin Pupil Assessment Program allows public and private schools to administer the same objective-referenced tests used for purposes of statewide assessment.

Overview of Statewide Tests

Schools may use the existing objective-referenced tests which have been developed for statewide assessment, instructions for administering the tests, test scoring arrangements, and reports of test results provided by the Wisconsin Pupil Assessment Program. This testing service offers several advantages to schools.

Advantages

- Large expenditures of teacher time for test development are eliminated. The tests are readily available for use.
- The tests, developed specifically for testing in Wisconsin, may be more relevant to the needs of Wisconsin schools than tests developed for national distribution.
- Schools may set performance standards for their students, basing the standards on the curriculum and expectations of individual schools or districts.
- State averages are available for objectives at grades four, eight, and twelve. Schools testing these grades may compare their students' scores with the state averages.

Type of Tests

All statewide tests offered within this service are objective-referenced. Objective-referenced tests are designed to measure performance in terms of specific objectives or skills. A list of objectives for all tests is found in this manual, beginning on page 4. The statewide tests are multiple-choice tests, with the exception of the writing tests, which require written samples.

Development

Objective-referenced tests used by the Wisconsin Pupil Assessment Program undergo a systematic and rigorous developmental process. Committees consisting of teachers, parents, employers, and others assume a central role in identifying the objectives or skills to be measured at various grade levels. Most test questions measuring selected objectives have been written by Wisconsin educators.

Content Areas

Tests have been developed in four areas: mathematics, reading, writing, and economic understanding. All of the statewide objective-referenced tests are designed to assess both academic and practical skills.

Although these tests were developed for the specific grade levels mentioned in this manual, they may be used at other grade levels. Schools should review items and objectives to determine at what grade level the tests can be most effectively used.

MATHEMATICS

A test measuring computation and its applications has been developed for use at the fourth grade. A second test assessing this skill area is available for the eighth and twelfth grades.

GEOMETRY

The test developed for fourth grade measures both geometry and number sentence skills. A second test, assessing geometry skills, has been developed for use at both eighth and twelfth grades.

READING

A separate test consisting of two parts, Educational Reading and Everyday Reading, has been developed for each of three grade levels: fourth, eighth, and twelfth.

WRITING

The fifth and eighth grade tests each contain two exercises: a narrative essay and an expository essay. The eleventh grade test contains three exercises: a narrative essay, a persuasive essay, and a job application letter. All exercises require written responses and are hand scored.

Both the holistic and primary trait scoring methods will be used to score the expository and the narrative compositions. The persuasive essay and business letter will be scored using only the primary trait method. These methods of hand scoring writing samples are described here:

Holistic Scoring

Holistic scoring is a method of hand-scoring essays to determine their relative quality. The term is derived from this method's emphasis on a reader's response to the whole essay--how well the student communicates. Readers are thoroughly trained to rate papers consistently according to pre-selected example essays which represent each score point.

Primary-trait Scoring

Primary-trait scoring determines if the essay successfully accomplishes the main purpose of a specified writing task.

The term is based on the assumption that particular writing tasks require particular approaches and no others if they are to be successful. The approach used by the writer to reach and affect the audience will be the most important (the primary) trait of a piece of writing. For example, the primary trait of a persuasive essay would be a systematic presentation of logical arguments directed to a specified audience. Successful papers will have that trait; unsuccessful papers will not, regardless of how clever or well-written they may be in other respects. Readers rate each essay according to predetermined criteria.

ECONOMIC UNDERSTANDING

A test of economic understanding is available for grade twelve. The test consists of two parts: Consumer Economics and Traditional Economics.

Table 1: Statewide Tests

The table below lists all test forms, showing the number of objectives, number of test items, and estimated administration time.

Grade Level*	Test Form	Objectives	Items	Estimated Time Needed to Complete Test Form**
4	Computation/Application	18	72	45-50 minutes
4	Geometry and Number Sentences	13	49	45-50 minutes
4	Educational Reading	4	35	45-50 minutes
4	Everyday Reading	9	36	45-50 minutes
5	Writing: Narrative Essay	1	1	30-45 minutes
5	Writing: Expository Essay	1	1	30-45 minutes
8	Computation/Application	23	90	45-50 minutes
8	Geometry	16	62	45-50 minutes
8	Educational Reading	9	36	45-50 minutes
8	Everyday Reading	11	44	45-50 minutes
8	Writing: Narrative Essay	1	1	30-45 minutes
8	Writing: Expository Essay	1	1	30-45 minutes
11	Writing: Narrative Essay	1	1	30-45 minutes
11	Writing: Job Application Letter	1	1	30-45 minutes
11	Writing: Persuasive Essay	1	1	45-50 minutes
12	Computation/Application	23	90	45-50 minutes
12	Geometry	16	62	45-50 minutes
12	Educational Reading	15	60	45-50 minutes
12	Everyday Reading	9	36	45-50 minutes
12	Consumer Economics	9	36	20-30 minutes
12	Traditional Economics	12	48	45-50 minutes

*Although all tests are identified with specific grade levels, they can be administered to students at other grade levels. Schools should carefully review test forms to determine suitability.

**Although students should be given as much time as needed to complete a test, nearly all students will finish within one class period (45-50 minutes).

TEST OBJECTIVES IN MATHEMATICS

Fourth Grade Mathematics Test: Computation and Applications

		% Students Mastering Objective*	STATE AVERAGE - Percent Correct**
COMPUTATION			
1.	Add whole numbers without regrouping.	86%	86%
2.	Add whole numbers with regrouping.	92%	89%
3.	Subtract whole numbers without regrouping.	87%	87%
4.	Subtract whole numbers with regrouping.	73%	77%
5.	Multiply whole numbers with basic facts.	95%	93%
6.	Multiply whole numbers beyond basic facts, no regrouping.	81%	83%
7.	Multiply whole numbers beyond basic facts, with regrouping.	75%	81%
8.	Divide whole numbers using ÷ symbol with basic facts.	88%	87%
9.	Divide whole numbers using)‾‾ symbol with basic facts.	86%	86%
10.	Divide whole numbers using the division algorithm, without regrouping or remainder.	63%	71%
MONEY APPLICATIONS			
11.	Solve simple, pictorial problems to find total costs and total coins.	86%	85%
12.	Solve multi-step pictorial problems to find total cost, tax, change, or earnings.	62%	68%
13.	Solve simple problems presented without a picture to find total cost and change from a purchase.	83%	83%
14.	Solve multi-step problems presented without a picture to find total cost, earnings, or change.	40%	53%
MEASUREMENT APPLICATIONS			
15.	Solve simple pictorial problems to find total weight or perimeter, or read a graph or ruler.	92%	91%
16.	Solve multi-step pictorial problems to find total miles, area, and differences in temperature.	72%	74%
17.	Solve simple problems presented without pictures to find total time, length, or capacity.	67%	72%
18.	Solve multi-step problems presented without pictures to find total time, units, or weight.	41%	55%

*The percent of students mastering the objective is based on correctly answering 75% of the questions measuring the objectives.

**Percentages are based on 1980 results for mathematics and reading. Results in geometry and economic understanding are for 1979.

Eighth and Twelfth Grade Mathematics Test: Computation and Applications

	% Students Mastering Objective		STATE AVERAGE - Percent Correct	
COMPUTATION	8	12	8	12
1. Add and subtract whole numbers.	85%	90%	85%	90%
2. Multiply whole numbers.	83%	87%	84%	87%
3. Divide whole numbers.	83%	86%	83%	86%
4. Find percents of whole numbers.	43%	63%	53%	68%
5. Find the percent one whole number is of another.	55%	71%	63%	75%
6. Find percent-fraction-decimal equivalents.	61%	73%	67%	75%
7. Add terminating decimals.	82%	89%	82%	87%
8. Subtract terminating decimals with regrouping.	72%	77%	76%	81%
9. Multiply decimals with regrouping.	76%	86%	78%	86%
10. Divide decimals with terminating quotients.	70%	73%	74%	77%
11. Use ratio and proration to solve equivalent fractions.	70%	80%	74%	82%
12. Add and subtract fractions with like denominators.	67%	80%	73%	83%
13. Add and subtract fractions with unlike denominators.	64%	71%	69%	74%
14. Multiply fractions using a horizontal form.	62%	70%	71%	75%
15. Divide using fractions.	58%	67%	66%	72%
MONEY APPLICATIONS				
16. Solve problems related to purchases (cost, possible purchase, change, discount, tax, or surcharge).	76%	89%	77%	86%
17. Solve problems dealing with earnings and savings.	76%	89%	78%	90%
18. Solve problems dealing with basic living and/or traveling expenses.	62%	83%	67%	82%
MEASUREMENT APPLICATIONS				
19. Solve problems dealing with length or area.	48%	72%	58%	74%
20. Solve problems dealing with weight and/or volume.	51%	76%	60%	77%
21. Solve simple and complex problems dealing with rate, time, and/or temperature.	35%	58%	50%	66%
22. Read graphs and maps.	83%	91%	83%	90%
23. Read and interpret tables.	56%	73%	73%	85%

TEST OBJECTIVES IN GEOMETRY

Fourth Grade Geometry Test	STATE AVERAGE - Percent Correct*
1. SIZE AND SHAPE: Identify pictures of two-dimensional figures; classify and apply concepts of size and shape.	70%
2. SETS AND POINTS: Identify picture; classify and note relationships of sets and points.	--
3. SYMMETRY: Recognize, classify, and apply concepts of symmetry.	81%
4. CONGRUENCE: Identify congruent figures, and apply to real life situations.	94%
5. SIMILARITY: Identify similar figures, recognize the relationship between similar and congruent figures, and apply the concept to real world objects.	91%
6. COORDINATE SYSTEMS: Number points on lines of a coordinate plane, and apply concept to real life situation.	74%
7. LANGUAGE: Use the terms < (less than), > (greater than), and = (equal to).	78%
8. COMMUTATIVE PROPERTIES: Replace missing numbers in open sentences.	--
9. Replace missing numbers in open sentences.	--
10. Place +, -, x, or ÷ in the placeholder of a sentence.	84%
11. Solve sentences which use the special properties of whole numbers.	--
12. Solve number sentences with placeholders in different positions.	89%
13. Identify mathematical statements which describe realistic situations.	--
14. Solve simple sentences which represent physical situations.	--

*State average is for 1979. If no figures are shown, this indicates considerable revision of test questions since the 1979 test was administered. Note: there are no mastery levels for tests administered statewide in 1979 and not administered in 1980.

Eighth and Twelfth Grade Geometry Test	STATE AVERAGE - Percent Correct	
	8th Grade	12th Grade
SIZE AND SHAPE		
1. Identify drawings of solids, quadrilaterals, and other polygons.	79%	87%
2. Classify angles, triangles, quadrilaterals, and other polygons.	60%	70%
3. Deduce a one-step conclusion when given a circle, triangle, quadrilateral, solid, and/or other information.	61%	80%
POINTS AND SETS OF POINTS		
4. Identify a point, line segment, polygon, radius or diameter; classify lines as parallel, perpendicular, or intersecting.	83%	90%
5. Relationships: Identify supplementary, complementary, vertical, and adjacent angles; identify relationships between points, lines, and planes.	54%	70%
6. Use known properties to deduce conclusions about plane and solid figures; apply concepts about sets of points to realistic situations.	55%	74%
7. Identify figures with one line of symmetry.	76%	82%
8. Identify image points, figures with more than one line of symmetry, and symmetrical real life objects.	62%	77%
9. Identify congruent angles, polygons, and corresponding parts of plane figures.	72%	85%
10. Use properties to deduce conclusions about congruent triangles and polygons.	66%	75%
11. Identify similar triangles, quadrilaterals, and plane figures.	89%	95%
12. Use properties to deduce conclusions about similar triangles.	--	--
13. Apply notions of similarity to real life situations.	55%	72%
COORDINATE SYSTEMS AND GRAPHS		
14. Use ordered pairs of whole numbers to identify a point, line, triangle, or quadrilateral.	65%	82%
15. Identify the coordinates and/or the points which fit a given description.	72%	87%

TEST OBJECTIVES IN READING

Fourth Grade Reading Test:

EDUCATIONAL READING	% Students Mastering Objective	STATE AVERAGE - Percent Correct
From examples of materials typically read by fourth graders, the student will be able to . . .		
1. answer questions about factual information in an expository selection.	64%	76%
2. answer questions about a poem.	32%	62%
3. answer questions about characterizations in a narrative selection.	57%	71%
4. answer questions about cause and effect in a passage.	71%	81%
5. identify the main idea in a passage.	41%	66%
EVERYDAY READING		
6. recognize appropriate information from a set of rules.	88%	88%
7. recognize appropriate information from a set of directions.	83%	82%
8. recognize appropriate information from a health warning.	75%	75%
9. recognize appropriate information from a telephone directory page.	90%	90%
10. recognize appropriate information from a dictionary page.	80%	81%
11. recognize appropriate information from a table of contents or index from a textbook.	69%	74%
12. recognize appropriate information from common street and highway signs.	88%	83%
13. recognize appropriate information from a highway map.	81%	80%
14. recognize appropriate information from a graph or chart.	72%	73%

Eighth Grade Reading Test:

	% Students Mastering Objective	STATE AVERAGE - Percent Correct
EDUCATIONAL READING		
From examples of materials typically read by eighth graders, the student will be able to . . .		
1. answer questions about factual information in an expository selection.	80%	80%
2. answer questions about the cause of events or actions in a newspaper or magazine selection.	88%	87%
3. answer questions about a poem.	66%	77%
4. identify the main idea in a newspaper or magazine selection.	58%	65%
5. identify the main idea in an expository selection.	59%	66%
6. infer a conclusion from an expository selection.	53%	63%
7. identify a writer's point of view in a newspaper editorial or magazine selection.	68%	72%
8. separate fact from opinion in a newspaper or magazine selection.	57%	65%
9. identify the cause and effect of an event described in a newspaper or magazine selection.	55%	64%
EVERYDAY READING		
10. recognize appropriate information from a set of guidelines.	96%	94%
11. recognize appropriate information from medical instructions.	84%	82%
12. recognize appropriate information from a health warning.	92%	89%
13. recognize appropriate information from a container label with instructions, warnings, or comparative buying information.	80%	80%
14. recognize appropriate information from a telephone directory page.	93%	91%
15. recognize appropriate information from a dictionary page.	80%	78%

	% Students Mastering Objective	STATE AVERAGE - Percent Correct
16. recognize appropriate information from a product warranty.	41%	55%
17. recognize appropriate information from a recipe.	76%	75%
18. recognize appropriate information from a business letter.	83%	79%
19. recognize appropriate information from a graph or chart.	54%	61%
20. recognize appropriate information from want ads or job descriptions.	83%	77%
21. separate fact from opinion in an advertisement.	61%	63%

Twelfth Grade Reading Test

EDUCATIONAL READING

From examples of materials typically read by twelfth graders, the student will be able to . . .

1. answer questions about factual information in a narrative selection.	78%	80%
2. answer questions about factual information in an expository selection.	84%	85%
3. answer questions about a poem.	82%	82%
4. answer questions about the cause of events or actions in a passage.	81%	85%
5. identify the main idea in a newspaper or magazine selection.	85%	83%
6. identify the main idea in a poem.	85%	84%
7. identify a generalization statement by making an inference from a passage.	77%	83%
8. answer questions about implied facts in a newspaper or magazine selection.	82%	83%
9. make an inference from a narrative selection.	82%	83%

	% Students Mastering Objective	STATE AVERAGE - Percent Correct
10. separate fact from opinion when reading a newspaper or magazine article such as an editorial.	79%	81%
11. identify bias in a newspaper or magazine selection.	88%	87%
EVERYDAY READING		
12. recognize appropriate information from operating instructions for an appliance.	92%	92%
13. recognize appropriate information from a business letter.	96%	94%
14. recognize appropriate information from a container label with instructions, warnings, or medical antidotes.	97%	96%
15. recognize appropriate information from a library catalog card.	76%	78%
16. recognize appropriate information from a dictionary page.	90%	86%
17. recognize appropriate information from a voting ballot.	82%	80%
18. recognize appropriate information from a computerized bill.	91%	88%
19. recognize appropriate information from want ads or job descriptions.	96%	92%
20. recognize appropriate comparative buying information from warranties, ratings, advertising, or container labels.	90%	89%

TEST OBJECTIVES IN WRITING

Fifth Grade Writing Test

The student will be able to . . .

1. write a story based on a photograph (narrative essay).
2. write a report given a short set of facts (expository essay).

Eighth Grade Writing Test

The student will be able to . . .

1. write a story based on a photograph (narrative essay).
2. write a report given a short set of facts (expository essay).

Eleventh Grade Writing Test

The student will be able to . . .

1. write a story based on a photograph (narrative essay).
2. write a job application letter.
3. write a persuasive essay.

TEST OBJECTIVES IN ECONOMIC UNDERSTANDING

Twelfth Grade Economic Understanding Test	STATE AVERAGE - Percent Correct*
CONSUMER ECONOMICS	
1. Students should understand how the consumer can affect what is produced in the American economic system.	88%
2. Students should understand the concept of insurance in terms of financial planning and risk management (e.g., health, social, life, liability, and property insurance).	--
3. Students should understand the concept of credit, including types and costs.	--
4. Students should understand the benefits and risks of various types of savings and investments.	80%
5. Students shall evaluate alternative choices when given a set of circumstances relating to needs and want (e.g., food, clothing, shelter, transportation, recreation, etc.).	75%
6. Students shall identify and be able to use a variety of sources of information about products and services (e.g., catalogs, consumer guides, institutions, yellow pages, newspapers, and personal contracts).	76%
7. Students shall interpret and evaluate advertising (e.g., amount of real savings, appeal to pressure tactics, representation of products and savings, etc.).	89%
8. Students shall be able to use information in selecting goods and services (e.g., packaging, labeling, guarantees, warranties, etc.).	73%
9. Students shall identify a financial plan for spending and saving that is comparable with one's personal values, goals, and resources.	79%
TRADITIONAL ECONOMICS	
10. Students should understand that because we have limited resources and the potential for unlimited wants there is a need to make choices and to economize.	65%
11. Students should understand the concepts of supply and demand.	69%

*The 1979 test was revised considerably in 1980. If no state averages are indicated, this indicates substantial change in test items.

12. Students should understand the concept of collective bargaining. 64%

13. Students should understand the concept of inflation. --

14. Students should understand the American economic system and how it compares with other economic systems. 59%

15. Students should understand the effect that current consumption patterns may have on the environment and future availability and utilization of resources - both for the individual and for the larger society. --

16. Students should understand the causes and effects of international trade. --

17. Students should understand the role of tax (fiscal) and monetary (money) policies in relationship to employment and price stability. --

18. Students should understand that income, including profits and wages, is generated through the production and exchange of goods and services. --

19. Students should understand that the government modifies the market economy by intervention and regulation. --

20. Students should be able to interpret economic data, as found in graphs, tables, ratios, index numbers, etc. --

21. Students should understand that specialization increases efficiency and leads to interdependence. 72%

General Testing Information

Sampling or Testing All Students

Whether a school chooses to test all students or select a random sample is dependent on at least two factors. One consideration should be the number of students in the "population" group (for example, the fourth grade). If the number is small, the required sample size will need to be nearly the same as the "population". In this case, there would be little to gain by sampling. For example, if the number of students in a grade is 60, a sample size of 52 would be needed. Table 2, indicating the sample size requirements, is shown on the next page.

A second consideration is whether a school wishes to obtain information about every student or whether a general group profile is desired. If the need is for information about the group as a whole, and the number of students is fairly large (greater than 150), a sample may be the best approach. Procedures for selecting random samples can be found in most introductory statistics books.

Setting Performance Standards

NOTE: THE CURRENT SCORING PACKAGE ALLOWS SCHOOLS TO SET A PERFORMANCE STANDARD FOR EACH TEST OBJECTIVE. THIS STANDARD IS LABELED "MASTERY".

It is important to recognize that you have the option of setting a performance standard for each test objective. However, should you choose not to set a standard, the standard will automatically be set at 75% (three of four items measuring an objective). This is necessary because a performance standard must be specified in order to complete the basic reports.

There are at least two ways to define this performance standard:

1. The standard may be set relatively low and be viewed as an "Adequate" level of performance.
2. The standard may be set relatively high and be viewed as a "Superior" level of performance.

A first step in setting a performance standard is to reach agreement about the level of learning expected of students. As indicated under Points 1 and 2 above, the standard can be established at two possible levels. DO NOT ASSUME THAT BECAUSE THE SCORING PACKAGE USES THE TERM "MASTERY" THAT THE STANDARD MUST BE DEFINED AT THE HIGHEST LEVEL OF LEARNING.

Shown in the following pages are two basic ways to set performance standards, labeled Methods I and II. In addition, there are brief summaries of two other methods which might be used to set performance standards. These are labeled Methods III and IV. Unlike the first two methods, which indicate how to set standards for specific objectives, Methods III and IV show how to set performance standards for the entire test.

Table 2
DETERMINING SAMPLE SIZE FROM A GIVEN POPULATION SIZE
N=Population Size, s=sample size

N	s	N	s	N	s	N	s
10	10	150	106	460	210	2200	327
15	14	160	113	480	214	2400	331
20	19	170	118	500	217	2600	335
25	24	180	123	550	228	2800	338
30	28	190	127	600	234	3000	341
35	32	200	133	650	242	3500	346
40	36	210	136	700	248	4000	351
45	40	220	140	750	254	4500	354
50	44	230	144	800	260	5000	357
55	48	240	148	850	265	6000	361
60	52	250	152	900	269	7000	364
65	56	260	155	950	274	8000	367
70	59	270	159	1000	278	9000	368
75	63	280	162	1100	285	10000	370
80	66	290	165	1200	291	15000	375
85	70	300	169	1300	297	20000	377
90	73	320	175	1400	302	30000	379
95	76	340	181	1500	306	40000	380
100	80	360	186	1600	310	50000	381
110	86	380	191	1700	313	75000	382
120	92	400	196	1800	317	100000	384
130	97	420	201	1900	320		
140	103	440	205	2000	322		

From Krijcie, R.V., and Morgan, D.W., "Determining Sample Size for Research Activities", Educational and Psychological Measurement, 1970, 30, pp. 607-610.

WHAT IS TO BE GAINED FROM SETTING PERFORMANCE STANDARDS PRIOR TO OBTAINING TEST RESULTS?

Some schools may choose to set performance standards after they have received test results for their students, basing the standards on the average scores and additional information about student performance. The alternative, establishing an *a priori* performance level, is advantageous for at least two reasons:

1. One reason is that knowledge of test results may bias expectations. For example, if students do exceptionally well on a given objective, a school might decide to raise its performance level. In contrast, lower scores may cause a lowering of expectations.
2. A second reason for setting performance standards prior to obtaining test results lies in the process itself. A conscientious and defensible process of setting performance standards requires a careful examination and evaluation of each test item. Awareness of student results may tempt one to be less thorough in the examination and evaluation process, and instead to rely primarily on student scores as the basis for setting performance standards.

WHAT SORTS OF GENERAL CONSIDERATIONS NEED TO BE MADE IN SETTING PERFORMANCE STANDARDS?

At the outset, it needs to be emphasized that there is no generally accepted procedure which allows one to set performance standards with ease. Each requires one to make hard decisions. What the various procedures do, however, is to provide guidelines for the setting of standards.

In addition to the above, there are other points to consider:

1. Items differ in their level of difficulty. This means that each item must be looked at in the process of setting standards.
2. Performance standards should be set at realistic levels, based on an understanding and awareness of school curricula and student experiences.
3. The process of setting performance standards is not something which is done once and then forgotten. Periodically, standards should be reviewed and adjusted.
4. Performance standards may be used to gather information on specific courses or programs. In other cases, they may be used for graduation requirements or to identify students who require remediation. If the performance standards are to be used for either of the latter two purposes, and especially graduation, it is imperative that the process be carefully planned and implemented. It is recommended that in these latter cases a large number of persons be involved.

SELECTING A COMMITTEE

Each of the methods of setting performance standards utilizes a committee of persons who are qualified to evaluate student capabilities. The selection of the committee involves two steps:

1. Reaching a decision about the size of the committee.
2. Reaching a decision about the composition of the committee.

The committee need not only consist exclusively of educators; schools may decide to include community leaders, school board members, etc. In terms of the size of the committee, the goal should be to avoid a group which is either too small or too large. If the committee is too small, it may be unrepresentative. If it is too large, it may be inefficient.

A fairly large group of perhaps twenty or more persons, which is representative of your community, may be the ideal. A large group which is diverse in its membership and viewpoints will be more credible to both educators and non-educators in your community. However, practical considerations may preclude the selection of a large and varied group of individuals. A reasonable compromise would be 10 to 15 persons.

CHOOSING THE OBJECTIVES WHICH ARE TO HAVE STANDARDS

After reaching a decision about the size and nature of the committee, there is a second necessary step. Schools must decide which objectives are to have performance standards. One way to reach this decision is to ask the following questions of each test objective:

1. Is this objective taught, and is it one our students should know?

2. Is this objective one which is <u>not</u> taught, but one which students should know?

3. Is this objective one which is <u>not</u> taught, and one which students should <u>not</u> be expected to know?

This second step is best done by people familiar with the school curriculum; however, non-educators can assume a role in this part of the process.

After careful review, a school may decide to set performance standards for all objectives. However, a reasonable approach would be <u>not</u> to set a performance standard for an objective if it is one which is not taught, and one which students should not be expected to know (point 3 above).

In this case, it is important to recall that should you decide not to set a performance standard for an objective, this will be automatically set at 75 percent by the computer. <u>This means that you must indicate a "Mastery" level of "Zero" items on the materials calling for identification of a performance standard.</u> Materials for setting performance standards will be sent with your test booklets and answer sheets and should be returned to Westinghouse DataScore after you have completed them.

METHOD I

The first method of setting performance standards consists of five steps:

1. Select a committee.
2. Have members of the committee assign each item (question) measuring an objective a numerical value. This value indicates how difficult they feel the item is. IN THIS STEP COMMITTEE MEMBERS MAY BE ASSIGNED DIFFERENT TASKS; IT IS NOT NECESSARY THAT THEY ALL DO THE SAME THING. For example, members may review different items and objectives.
3. Have members of the committee discuss their ratings with the others. On the basis of the discussion individuals may choose to change their ratings of items. A dialogue among committee members is an important part of the process.
4. Tally the ratings of the committee members and determine the average Level of Difficulty for all items measuring each objective.
5. Set the performance standard(s).

After a decision has been reached about the size and composition of the committee and the objectives which are to have performance standards, the next step is to have each committee member assign a numerical value to the test items, indicating the Level of Difficulty. A scale of recommended values is given below:

The Level of Difficulty

Value	Description
1	A very difficult item; perhaps less than 25% of our students should be expected to answer this item correctly.
2	A moderately difficult item; perhaps 25%-50% of our students should be expected to answer this item correctly.
3	A moderately easy item; perhaps 50%-75% of our students should be expected to answer this item correctly.
4	A very easy item; perhaps 75% or more of our students should be expected to answer this item correctly.

EXAMPLE OF METHOD I

An example of how Method I might be used follows.

Items 35, 36, 37, and 38 are measures of Objective 15. The Level of Difficulty, as judged by one of the committee members, is shown to the right of each item.

EXAMPLE:

Objective Number 15

Item Numbers	Level of Difficulty
35	4
36	1
37	2
38	2
	Total = 9

Average for Objective = 2.25

In this instance, item 35 has been judged as "very easy", item 36 as "very difficult", and items 37 and 38 as "moderately difficult". The average of the four items is 2.25 (9 divided by 4).

Suppose that the committee consists of ten members and that their average ratings for Objectives 15 are as follows:

Committee Member	Average Level of Difficulty (Objective 15)
Individual 1	2.25
Individual 2	2.50
Individual 3	2.00
Individual 4	2.25
Individual 5	1.75
Individual 6	2.25
Individual 7	2.00
Individual 8	1.75
Individual 9	2.25
Individual 10	2.00
	Total = 21.00

Average for Committee = 2.10

Based on the average ratings of the ten committee members, Objective 15 has a Level of Difficulty of 2.10. The table below shows recommended performance standards based on the average Level of Difficulty as calculated in this example.

The value 2.10 is sufficiently close to the value "2" so that reasonable performance levels for "Adequate" and "Superior" would be 2 and 3, respectively.

	Recommended Performance Levels		
	Average Level of Difficulty	Adequate	Superior
Very Difficult	1	1	2
Moderately Difficult	2	2	3
Moderately Easy	3	3	4
Very Easy	4	4	4

Problems occur when the average Level of Difficulty approximates the values, 1.5, 2.5, or 3.5. For example, if the average Level of Difficulty is 1.55., the committee members must decide whether to set the performance levels for "Adequate" and "Superior" at 1 and 2, respectively, or possibly 2 and 3.

METHOD II

The second method of setting performance standards is a slight modification of one offered by Nedelsky in 1954*. Like Method I, Method II also consists of a six-step process designed to provide guidelines for setting performance standards. The five steps are as follows:

1. Select a committee.
2. Have members of the committee assign each item (question) measuring an objective a numerical value. This value is based on the number of answers that an average student should be able to recognize as WRONG. IN THIS STEP COMMITTEE MEMBERS MAY BE ASSIGNED DIFFERENT TASKS: IT IS NOT NECESSARY THAT THEY ALL DO THE SAME THING. For example, members may review different items and objectives.
3. Have members of the committee discuss their ratings with the others. On the basis of the discussion individuals may choose to change their ratings of items. A dialogue among committee members is an important part of the process.
4. Tally the ratings of the committee members to determine an average Probability Value for the items measuring an objective.
5. Set the performance standard(s).

EXAMPLE OF METHOD II

In the following example, there are four items measuring an objective: numbers 9, 10, 11, and 12. One of the committee members has reviewed the items and indicated the number of answers for each item that he thinks an average student should be able to eliminate as being obviously wrong.

In the case of item 9, the number "1" has been circled, indicating that it is the opinion of this committee member that an average student should be knowledgeable enough to eliminate at least one answer out of four as being possible. This means that even if the student is forced to guess the correct answer to the question, there is a one in three chance of guessing correctly. For this reason, the figure .33 is entered in the table. This value is referred to as the Probability Value.

Item Number	Number of Answers an Average Student Should Recognize as Wrong				Probability of Randomly Guessing the Correct Answers
9	0	(1)	2	3	.33
10	(0)	1	2	3	.25
11	0	1	(2)	3	.50
12	0	1	2	(3)	1.00
				Individual's Total Rating =	2.08

Item 10 is thought to be very difficult. In the opinion of the committee member, an average student would be unable to eliminate any of the answers as being obviously wrong. As a result, the student has only a one in four change of guessing the correct answer. A one in four chance provides a Probability Value of .25.

---

*Nedelsky, L. "Absolute Grading Standards for Objective Tests", Educational and Psychological Measurement, 1954, 12, pp. 3-19.

For item 11 the value is .50, while item 12 has a value of 1.00. Item 12 is felt to be very easy, and the number "3" is circled. Because three of the answers should be recognized as wrong or "absurd", the student is left with only one reasonable choice. Thus, the Probability Value is 1.00.

Suppose that ten committee members are used to judge the four items, with the following results:

Committee Member	Individual's Total Ratings
Individual 1	2.08
Individual 2	2.00
Individual 3	2.00
Individual 4	1.99
Individual 5	1.83
Individual 6	1.83
Individual 7	2.00
Individual 8	1.83
Individual 9	1.96
Individual 10	2.00
	Total = 19.52

Average for Objective = 1.95

For the ten committee members the average Probability Rating for the objective is 1.95, which can be rounded to 2.00. The table below offers recommended performance levels, based on the average of the committee members for an objective measured by four items

	Recommended Performance Levels		
	Average Level of Difficulty	Adequate	Superior
Very Difficult	1	1	2
Moderately Difficult	2	2	3
Moderately Easy	3	3	4
Very Easy	4	4	4

Items measuring an objective having a Probability Rating close to "1" are very difficult items, and the "Adequate" and "Superior" levels should be quite low. The recommended values are 1 and 2. For the example given above, the average Probability Rating of the committee is nearly 2. Thus, it would be reasonable to set the "Adequate" level at 2 and the "Superior" level at 3.

As with Method I, the most difficult decisions arise when the ratings approximate 1.5, 2.5, or 3.5. Members of the committee need to decide whether to round up or down.

TWO STANDARD SETTING METHODS FOR AN ENTIRE TEST

There are two additional methods which can be used to set performance standards. These methods differ from those presented earlier in this testing services manual, in that they represent procedures to set standards for the test in its entirety, rather than for each objective on the test. The procedures described earlier (Methods I and II) are designed to set performance levels which are to be returned with the answer sheets. The computer scoring system is programmed to report how well students have met predetermined performance levels.

METHOD III

Method III is similar to Method I, in which committee members are asked to assign a value to test items which reflects how difficult they feel each item is. The recommended scale of values is shown on page 19. In the following example, a test consists of 24 items, and each item has received a Level of Difficulty score.

Item Number	Level of Difficulty	Item Number	Level of Difficulty	Item Number	Level of Difficulty
1	1	9	4	17	3
2	3	10	2	18	4
3	4	11	3	19	2
4	2	12	3	20	1
5	3	13	1	21	4
6	2	14	2	22	2
7	3	15	3	23	2
8	2	16	2	24	3

Total = 61

Average Level of Difficulty = 2.54 (61 ÷ 24)

During the process of judging the difficulty level of test items, committee members may discuss their decisions and change them. After this step is completed, there are five additional steps needed to develop estimates for performance standards.

Step 1 The difficulty values of each item are totaled. In this example, the total is 61.

Step 2 The total is divided by the number of items on the test to obtain an average Level of Difficulty. In this case, the difficulty value is 2.54: 61 ÷ 24 items = 2.54.

Step 3 The average Level of Difficulty is divided by 4.0 (the top value on the four point Level of Difficulty scale). The result is a percentage:

$$\frac{2.54}{4.00} = 64\%$$

Step 4 The number of items on the test is then multiplied by the percentage obtained in Step 3. The resulting average performance score is 15.36: .64 x 24 items = 15.36.

Step 5 The average performance scores, determined for each committee member, are then averaged. In the following example, the average is 15.71. This value represents the number of items the committee expects the average student to answer correctly.

Committee Member	Average Performance Score
Individual 1	15.36
Individual 2	17.65
Individual 3	16.90
Individual 4	15.66
Individual 5	14.50
Individual 6	15.00
Individual 7	15.48
Individual 8	17.00
Individual 9	14.56
Individual 10	15.00
	Total = 157.11

Average for the Committee = 15.71

The value, 15.71, should be rounded to 16 items. In effect, this value represents an average Level of Difficulty for the test as a whole and would represent a reasonable choice for an "Adequate" level. Should a school wish to set a higher "Mastery" level, committee members would need to select a higher score. In this example, a "Superior" level would be somewhere between 17 and 24.

TWO STANDARD SETTING METHODS FOR AN ENTIRE TEST

There are two additional methods which can be used to set performance standards. These methods differ from those presented earlier in this testing services manual, in that they represent procedures to set standards for the test in its entirety, rather than for each objective on the test. The procedures described earlier (Methods I and II) are designed to set performance levels which are to be returned with the answer sheets. The computer scoring system is programmed to report how well students have met predetermined performance levels.

METHOD III

Method III is similar to Method I, in which committee members are asked to assign a value to test items which reflects how difficult they feel each item is. The recommended scale of values is shown on page 19. In the following example, a test consists of 24 items, and each item has received a Level of Difficulty score.

Item Number	Level of Difficulty	Item Number	Level of Difficulty	Item Number	Level of Difficulty
1	1	9	4	17	3
2	3	10	2	18	4
3	4	11	3	19	2
4	2	12	3	20	1
5	3	13	1	21	4
6	2	14	2	22	2
7	3	15	3	23	2
8	2	16	2	24	3

Total = 61

Average Level of Difficulty = 2.54 (61 ÷ 24)

During the process of judging the difficulty level of test items, committee members may discuss their decisions and change them. After this step is completed, there are five additional steps needed to develop estimates for performance standards.

Step 1 The difficulty values of each item are totaled. In this example, the total is 61.

Step 2 The total is divided by the number of items on the test to obtain an average Level of Difficulty. In this case, the difficulty value is 2.54: 61 ÷ 24 items = 2.54.

Step 3 The average Level of Difficulty is divided by 4.0 (the top value on the four point Level of Difficulty scale). The result is a percentage:

$$\frac{2.54}{4.00} = 64\%$$

Step 4 The number of items on the test is then multiplied by the percentage obtained in Step 3. The resulting average performance score is 15.36: .64 x 24 items = 15.36.

Step 5 The average performance scores, determined for each committee member, are then averaged. In the following example, the average is 15.71. This value represents the number of items the committee expects the average student to answer correctly.

Committee Member	Average Performance Score
Individual 1	15.36
Individual 2	17.65
Individual 3	16.90
Individual 4	15.66
Individual 5	14.50
Individual 6	15.00
Individual 7	15.48
Individual 8	17.00
Individual 9	14.56
Individual 10	15.00
	Total = 157.11

Average for the Committee = 15.71

The value, 15.71, should be rounded to 16 items. In effect, this value represents an average Level of Difficulty for the test as a whole and would represent a reasonable choice for an "Adequate" level. Should a school wish to set a higher "Mastery" level, committee members would need to select a higher score. In this example, a "Superior" level would be somewhere between 17 and 24.

METHOD IV

Method IV is similar to Method II. Each member of the committee is asked to judge the probability of an "average" student randomly guessing the correct answer. For each item a committee member is asked to indicate the number of answers that an average student should be able to eliminate as possible answers. The calculated value represents the probability that a student would randomly guess the correct answer.

An example is given for a test consisting of 32 items.

Item Number	Probability Value	Item Number	Probability Value	Item Number	Probability Value
1	.25	12	.25	23	.50
2	.33	13	.25	24	1.00
3	.50	14	.50	25	.33
4	.33	15	.33	26	.25
5	.50	16	.50	27	.25
6	1.00	17	1.00	28	.50
7	.25	18	.25	29	.33
8	.33	19	.33	30	.33
9	.50	20	.25	31	.50
10	.33	21	.50	32	.25
11	.50	22	.33		

Individual's Total Rating = 13.55

It is obvious that in the opinion of this committee member the test is very difficult. After an average student has eliminated the "obviously wrong" answer choices, he or she would have a high probability of guessing the correct answers to only 13 or 14 items out of the 32.

The next step is to determine the average Total Rating of all committee members. In the example blow, the average is 13.28.

Committee Member	Total of Probability Value
Individual 1	13.55
Individual 2	12.90
Individual 3	14.15
Individual 4	12.05
Individual 5	13.00
Individual 6	14.50
Individual 7	12.75
Individual 8	12.80
Individual 9	13.70
Individual 10	13,40

Total = 132.80
Average for Committee = 13.28

The average rating determined by the ten committee members should be rounded to 13 items. This value could serve as the "Adequate" level for the 32 item test. Determination of a "Superior" level could be set as a result of a discussion among committee members. An alternative would be to substitute the word "exceptional" for "average" student and repeat the process outlined above.

A Concluding Remark Concerning Performance Standards

The methods outlined in this manual represent examples of ways to establish performance standards. Schools may, of course, modify them or develop their own procedures. Regardless of the direction taken, it is critical that the process of setting performance standards be systematic and thorough. Generally, the process itself will have more credibility if large numbers of persons are involved. Though all the examples of committees in this paper consist of ten persons, this committee size represents a minimum.